

v1.1 | 2024

AI AUDITING LITEPAPER

 auditone.io

 [@auditone_dao](https://twitter.com/auditone_dao)

TABLE OF CONTENTS

<i>Executive summary</i>	02
<i>Introduction</i>	02
<i>Current state of AI systems auditing</i>	03
<i>About the framework</i>	05
Security	07
Technical Assessment	07
Regulatory Compliance	07
Ethics	08
Data Governance	08
<i>AuditOne: Your Partner for AI Auditing</i>	09
Preliminary Analysis	10
Expert Analysis	10
Audit Results	11
Additional Support	11
<i>Case study: Miljn Audit</i>	12

01-Executive summary

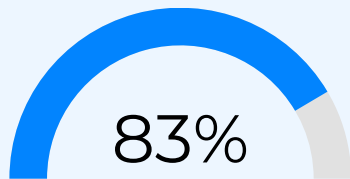
The proliferation of AI Systems has an immense potential to improve people's everyday lives. However, their rise has raised concerns about their fairness, security, and regulatory compliance, when using AI systems. Most AI professionals find it difficult to identify active biases in their AI systems, and there is a dire need for AI security. Regulators worldwide, foremost the EU have implemented strict regulatory frameworks around AI with severe penalties for non-compliance.

Current frameworks fall short and fail to assess the technical depth necessary to fully assess the functional components of AI systems comprehensively. AuditOne offers a **Security, Technical Assessment, Regulatory Compliance, Ethics, and Data Governance** framework called **STARED**. This approach assesses AI systems across multiple categories to ensure they fulfill ethical, technological, and legal criteria while also promoting safety, equality, and transparency. Our framework would be targeted towards AI or Machine Learning development teams seeking robust auditing solutions, as well as businesses that are deploying AI systems and preparing for certifications under the emerging EU AI Act framework.

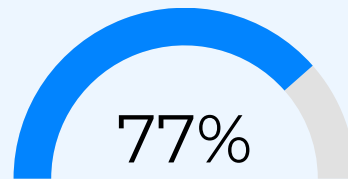
02-Introduction

Artificial Intelligence (AI) is increasingly integrated into various sectors, offering transformative benefits. However, with its rapid expansion come critical concerns including Ethics, Security, and Regulatory oversight. One major concern is fairness, with 83% of machine learning professionals saying it's hard to identify bias in AI systems. Security also remains a major issue, as 77% of businesses experienced AI breaches last year, according to [HiddenLayer's 2024 AI Threat Landscape Report](#), putting sensitive data at risk. On the regulatory front, new frameworks, like the EU's AI Act, propose fines of up to 7% of global revenue or 35 Million Euros, whichever is higher for non-compliance with AI standards stated in the EU AI Act under [Article 99: Penalties](#). More than 80% of AI projects fail, which is double the failure rate of non-AI information technology projects.

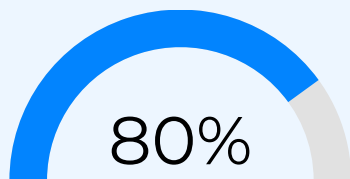
Key contributing factors include insufficient technical evaluation of model performance, inadequate infrastructure for data management, and a lack of quality data for training models (Ryseff, De Bruhl, & Newberry, [*The Root Causes of Failure for Artificial Intelligence Projects and How They Can Succeed*](#)).



ML professionals find it challenging to identify bias in AI Systems



Businesses faced AI breaches last year



Of AI projects fail, which is 2x the failure rate of traditional IT projects

€35 million

in fines

In Regulatory Risk, non-compliance with the EU AI Act could result in fines of up to 7% of global revenue or 35 million euros.

This is where AI Auditing becomes critical. AI systems auditing involves evaluating AI systems to make sure that they align with ethical, regulatory, and technical standards, thereby promoting trust and compliance. Through an independent assessment from experts in the field, AI auditing offers the opportunity to identify measures to improve AI systems in relevant areas.

03-Current state of AI systems auditing

This section explores the existing frameworks and regulations that guide AI auditing, offering an overview of how they address various aspects such as technical robustness, security, compliance, data governance, and ethics. The following table provides a comparison of the top frameworks, highlighting their focus areas, inclusions, and potential drawbacks.

Framework/ Regulations	Focus areas	What it includes	Drawbacks
NIST AI Risk Management Framework	Technical, Security, Compliance	Risk management for AI systems, including bias detection, fairness, security, and robustness. Emphasizes risk assessment and mitigation.	Limited ethical considerations, primarily focused on technical risk management.
ISO 42001	Security, Compliance	Guidelines for securing AI systems through information security management, focusing on data protection, privacy, and risk mitigation.	Limited focus on algorithmic fairness and ethical concerns, mostly focused on security.
GDPR	Data Governance, Privacy, Compliance	Ensures AI systems comply with data protection regulations, focusing on transparency, data rights, and lawful data processing.	Focused on privacy and data protection, not a comprehensive framework for technical or ethical auditing.
Ethics Guidelines for Trustworthy AI (AI HLEG)	Ethical, Compliance	Promotes fairness, transparency, accountability, and human oversight in AI, ensuring societal well-being and ethical development.	Non-binding guidelines with limited enforceability and technical implementation methods.
OECD AI Principles	Ethical, Transparency	Encourages Responsible AI development through principles like transparency, accountability, and fairness.	High-level principles with limited technical guidance and enforcement mechanisms.

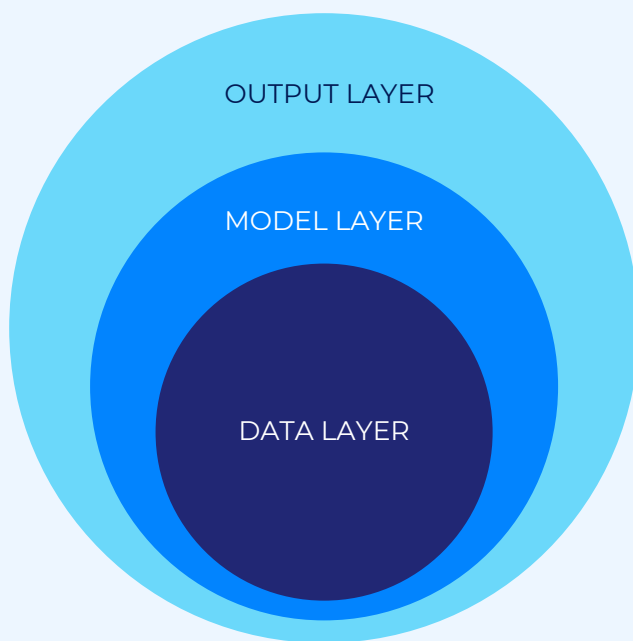
From the analysis, it can be concluded that most of the frameworks try to set high-level standards, focusing mainly on setting guidelines for security, ethics, and compliance. However, they often fall short when it comes to deeply examining the **functional components** of AI systems, such as Algorithmic Transparency, Bias Mitigation, Robustness, and Performance. Although these frameworks have the necessary principles and risk management approaches, they lack the technical depth to assess how AI models operate in real-world scenarios. This gap demands the need for more robust frameworks that go beyond surface-level evaluations to make sure that AI systems function as intended in all contexts.

To address this, **AuditOne** has developed the **STARED** framework, which stands for **S**ecurity, **T**echnical **A**ssessment, **R**egulatory Compliance, **E**thics, and **D**ata Governance. It aims to provide a 360-degree audit for organizations developing and using AI systems to ensure that all AI assessment verticals are covered to improve the AI system's performance and help them prepare for the upcoming regulations and certifications.

04-About the framework

A comprehensive audit requires a robust framework that assesses the system across multiple verticals; this is where the STARED framework comes into play. The auditing process involves evaluating different layers of AI systems, including the data, model, and output layers. Each of these layers affects the overall performance and accountability of the AI system.

The AI auditing process is divided into three primary layers:



- Evaluate accuracy, fairness, and ethical acceptability of AI Outputs
- Assesses model transparency, bias, technical soundness, and robustness
- Ensures data governance, privacy, and quality throughout the AI lifecycle.

Data Layer

This layer focuses on the data collected, how it is processed, and how it is managed within the system. Auditing this layer ensures that data governance, privacy, and quality are maintained throughout the AI lifecycle.

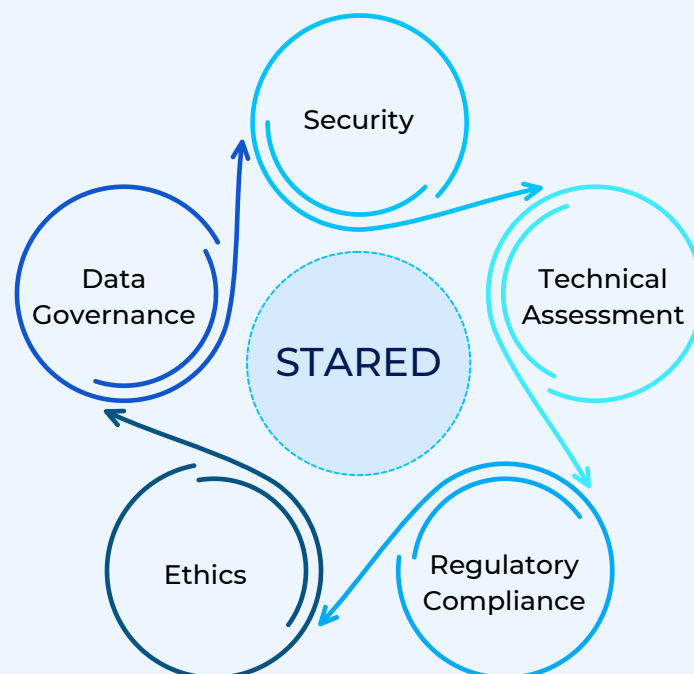
Model Layer

This layer includes the architecture, complexity, and overall functioning of the AI model itself. Auditing at this level includes technical assessment and security of the models to ensure that models are transparent, unbiased, technically sound, and robust.

Output Layer

At this layer, the focus is on the quality and fairness of the outputs generated by the AI system. Evaluating the output ensures that the predictions made in classical Machine Learning models or generated content are accurate, coherent, and ethically acceptable. This also includes the compliance of the user interface.

Below is a breakdown of each component of the STARED framework and key areas that are evaluated as part of the specific vertical. For detailed areas, refer to Appendix A.



Security

As AI systems must be protected from potential threats and vulnerabilities, security is given the top priority. In the STARED framework, Model Robustness is evaluated using standards such as the OWASP Top 10 for Machine Learning (ML) and Large Language Models (LLM). This ensures that the AI models are secured against adversarial attacks like model poisoning, sensitive information disclosure, prompt injection, and supply chain vulnerabilities. A security score is assigned to measure the AI system's robustness and identify any security vulnerabilities that help organizations fix them before it breaks the user's trust.

Technical Assessment

The technical aspect of AI auditing focuses on assessing the model and system architecture, risks associated, and monitoring methods. Under the STARED framework, Risk Assessment evaluates risks associated with both LLM-powered systems and classical machine learning models. It also includes an Architecture evaluation, which measures the complexity of the AI model, inference latency, and the performance of the model pipeline. Assessment related to monitoring systems to alert in case inappropriate content is generated and tracking model performance over time is included as a part of the Technical Assessment.

For LLM-powered AI systems, some of the technical metrics include the Drift Score, which detects shifts in language patterns, and the Output Quality, which measures the coherence and relevance of responses based on human ratings. Classical ML models are assessed based on prediction metrics like accuracy, recall, precision, and retraining frequency to make sure models maintain high accuracy and reliability over time.

Regulatory Compliance

Compliance with laws and regulations is another critical focus of the framework. AI systems must adhere to standards such as GDPR and the EU AI Act, especially in areas related to data protection and Ethics. Compliance-related Risk Management is a central aspect of quality management, as STARED assigns risk scores across different layers of AI systems, including data, models, and overall system operations.

AuditOne's main service in this area provides organizations with Guidance to assist them in achieving compliance. This approach ensures that organizations

are prepared to meet regulatory requirements while also addressing the risk management aspect, which is important for mitigating potential legal or operational issues.

Ethics

The FAT principle—Fairness, Accountability, and Transparency—is at the heart of AI Ethics. Under the STARED framework, this principle is applied across all the layers.

In the Data layer, Fairness is measured by assessing bias in pre-trained or fine-tuned models for LLMs and by examining class distribution parity for classical ML models. Accountability ensures that all data sources, preprocessing steps, and dataset changes over time are documented. Transparency requires the disclosure of data sources and any relevant statistics, making the data pipeline fully traceable.

In the Model layer, Fairness involves evaluating biases in model generation and decision-making processes. Accountability includes keeping logs of fine-tuning and tracing how specific decisions are made by the model.

Transparency focuses on the explainability of the model's output, especially in systems that use techniques like Retrieval-Augmented Generation (RAG) or fine-tuning in Generative AI applications and black box in classical ML systems.

In the Output layer, Fairness ensures there is no bias in the generated content. Accountability includes implementing mechanisms to handle errors in generated outputs for LLMs, and error-reporting processes for classical ML models. Transparency in this layer includes the provision of confidence scores in the AI-generated responses, ensuring that users can trust the output.

Data Governance

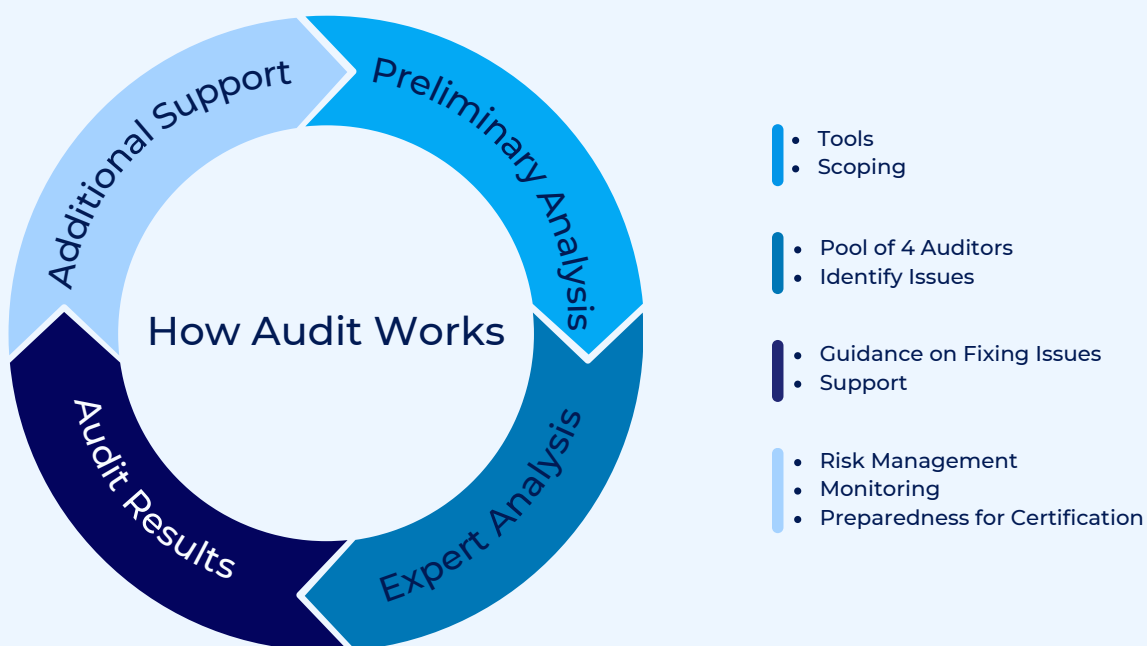
Data Governance focuses on assessing proper Access controls, Data privacy, Data spread, and Data quality. This ensures that data within AI systems is secure, properly managed, and that access is limited to authorized individuals. Data governance is important for maintaining the integrity of the data pipeline and ensuring compliance with privacy regulations. Effective data governance also ensures that data used for AI training and inference is high-quality, free from biases, and managed according to strict privacy standards.

05-AuditOne: Your Partner for AI Auditing

Experts from different areas like technology, cybersecurity, law, and ethics are needed to conduct a successful AI systems audit. However, organizations struggle to find and combine this expertise in one team to effectively develop and audit AI systems. Bringing together experts from diverse fields like AI/ML, Data Governance, Ethics, Security, Law, and industry is important for ensuring AI systems are technically sound and ethically aligned.

At AuditOne, our team of auditors has years of experience in these key areas, providing a comprehensive, interdisciplinary approach to AI system audits. This diverse expertise allows us to assess your AI systems thoroughly, and by identifying potential issues early, we help organizations address concerns before they escalate, ensuring that AI systems operate responsibly and meet regulatory and ethical standards. These audits can also provide the preparedness for upcoming certifications in AI like ISO 42001.

The procedure for conducting an audit at AuditOne is presented below:



Preliminary Analysis

AuditOne begins by conducting a preliminary analysis using specialized tools to assess the AI system type and risk level. This step also involves defining the scope of the audit and providing a customized audit plan that suits the complexity of the AI system and the organization's specific requirements.

Expert Analysis

At AuditOne, we create auditor pools with interdisciplinary experts to ensure a comprehensive evaluation of AI systems, helping organizations meet regulatory standards, mitigate risks, and enhance operational efficiency.

Meet the AI experts who helped us create this framework.

Tobias S. Potthoff

A lawyer with over 20 years of experience, Tobias holds an EMBA from St. Gallen, CIPM, and PECB certifications and is an expert in navigating international regulatory landscapes. His work at Deutsche Telekom included ensuring GDPR compliance and leading multi-national audit projects. A certified lead auditor in ISO/IEC 27701, Tobias ensures privacy frameworks are aligned with both business goals and legal requirements.

Hubert Jackowski

Hubert is a seasoned data privacy and compliance expert, specializing in global data protection regulations like GDPR. He is a professional member of International Compliance Association with advanced certificate in business compliance. As a Group Data Protection Officer, he implements strategic frameworks that align with business objectives while ensuring legal compliance. Certified in CIPP/E and CIPT, Hubert offers a unique perspective that turns regulatory challenges into competitive advantages.

Sivaram Annadurai

Sivaram is an AI Advisor and Strategist with over 20 years of experience, specializing in AI-powered products/solutions across image, video, text, and data. Sivaram has had an extensive career in product development and consulting roles for numerous high-profile organizations, including Amazon, Microsoft, Publicis Sapient, Sensormatic, Unilever, Tesco, Macy's, The Container Store, The Children's Place, and Belcorp. His consulting experience spans several domains, including Retail, Supply Chain, Beauty, Fashion, Cybersecurity, and Real Estate, utilizing GenAI/Vision/NLP/Deep Learning. He holds two granted patents in the Computer Vision Retail domain.

Alongside his work in industry, Sivaram has also made long-standing contributions to the field via his teaching, mentoring, design and development of AI curriculum, as well as his training and mentorship through organizations that include upGrad, Emeritus, Interview Kickstart, and the Institute of Product Leadership.

Gayathri Devi B R

Gayathri is an ISO/IEC 42001 certified Cybersecurity and AI Governance expert with over 15 years of experience, specializing in Responsible AI solutions, risk management, and compliance across various industries. She has an extensive career in advisory and management roles for a number of high-profile organizations, with notable examples such as Shell, Unilever, RBS, Wipro, and Infosys. Her expertise bridges several areas (including AI Governance, Information Security, Cloud Security, and Regulatory Compliance) using technologies such as GenAI, Cloud Platforms and IT Risk Management frameworks. An SMIEEE and Executive Committee Member of IEEE Bangalore Section, CIS, and WiE, she is a speaker at public forums and conferences, delivers expert lectures, and has developed cybersecurity and AI governance courses.

Audit Results

Following the audit, AuditOne delivers a comprehensive report that outlines the findings. This report goes beyond merely listing vulnerabilities, it provides detailed analysis, specific guidelines, and actionable recommendations aimed at enhancing the performance, robustness, security, and compliance of the AI system. Our insights cover all identified issues and offer solutions to address both immediate concerns and long-term improvements in system performance.

In addition to technical guidance, AuditOne assists in effectively communicating these security achievements to stakeholders. We publish the report and give a shoutout to the community via social media showcasing your commitment to security. This helps build trust with customers, investors, and partners, highlighting the positive outcomes of the audit and your proactive approach to maintaining a secure, compliant AI environment.

Additional Support

AuditOne also offers continuous monitoring of the AI system post-audit, ensuring ongoing risk management. Additionally, we assist organizations in preparing for future certifications, ensuring that their systems meet evolving industry standards and regulatory requirements.

06-Case Study: Miljn Audit

AuditOne conducted an AI audit of Miljn, an AI-driven platform aimed to transform how users engage with books. The audit was aimed to evaluate AI systems in multiple areas that included - security, technology, ethics, and compliance. Experts performed manual assessments, automated tool analysis, and data collection via questionnaires. The audit focused on Miljn's user-level functionality, data handling, and adherence to relevant regulations. Findings and recommendations were provided privately to Miljn, ensuring the platform's alignment with compliance and user impact standards. The summary of the Audit can be accessed through the following link: <https://www.auditone.io/audit-report/miljn-audit>

Conclusion

AI systems have become an important part of modern technology but pose challenges related to security, fairness, and regulatory compliance. Existing frameworks often fail to address the technical depth required for a thorough AI audit. AuditOne's STARED framework can fill this gap by assessing AI systems at the functional components level across Security, Technical Assessment, Regulatory Compliance, Ethics, and Data Governance. This comprehensive audit approach ensures that AI systems are secure, ethically aligned, and compliant with emerging AI regulations. AuditOne's STARED framework gives businesses confidence in their AI deployments while helping them prepare for future certifications and regulatory challenges.

07-Appendix A: Areas of Auditing

STARED	Topic
Security	Model Robustness (OWASP Top 10)
	Safe, Valid, Reliable, Resilient Models
	AI Security Architecture Review
	AI Risk & Threat Management
	AI Controls
	Decommissioning Policies
Technical Assessment	Models-related Risk Assessment
	Architecture Evaluation
	Prompt Engineering
	AI System Inventory
Regulatory Compliance	Trust AI Regulation
	EU AI Act
	AI Auditing & GDPR
	Monitoring & Incident Response
	Legal & Regulatory Compliance

Ethics

Accountability

Responsible AI

Explainable AI

Transparency

Privacy

Fairness, Bias & Discrimination

Transparency & Accountability

Data Governance

Architecture Evaluation

Data Collection

Data Lineage

Data Security

Data Ubiquity

Data Training

Data Protection

Data Privacy

Contact



auditone.io



[@auditone_dao](https://twitter.com/auditone_dao)



hello@auditone.io



The trust layer of our
multi-stakeholder world.